

# Parag Jain

Email: [jainparag.in@gmail.com](mailto:jainparag.in@gmail.com)

Web: [parajain.github.io](https://parajain.github.io)

Linkedin: [www.linkedin.com/in/paragjain1](https://www.linkedin.com/in/paragjain1)

Google Scholar: <https://shorturl.at/wizL2>

## Industry Experience

<b>Google DeepMind</b> Student Research Intern	July 2023 - Nov 2023
<b>IBM Research</b> Research Engineer	Aug 2016 - Sept 2019
<b>Amazon</b> Software Development Engineer	July 2015 - Aug 2016

## Education

<b>University of Edinburgh</b> CDT-NLP PhD, Natural Language Processing, Advisor: <a href="#">Prof. Mirella Lapata</a> Funded by Huawei and UKRI.	Sept 2019 - Current
<b>Indian Institute of Technology Hyderabad, India</b> M.Tech, Computer Science & Engg, Advisor: <a href="#">Prof. Vineeth N Balasubramanian</a> <b>Thesis:</b> Metric Learning for Clustering in Streaming Large-Scale Data - <a href="#">link</a> Awarded A+ (highest possible) for research thesis	July 2013 - June 2015 CGPA: 8.92/10
<b>Jabalpur Engineering College, India</b> B.E., Computer Science & Engg	July 2008 - June 2012 CGPA: 7.5/10

## Conference Tutorial

- [Storytelling from Structured Data and Knowledge Graphs : An NLG Perspective](#)  
Presented at the Annual Meeting of the *Association for Computational Linguistics* (ACL), 2019, Florence (Italy)

## Publications

### Journal

- **Jain, P.**, Lapata, M.  
[Memory-Based Semantic Parsing](#) - Transactions of the Association for Computational Linguistics (TACL) 2021
- Laha, A.\*, **Jain, P.\***, Mishra, A.\*, Sankaranarayanan, K.  
[Scalable Micro-planned Generation of Discourse from Structured Data](#) - (Computational Linguistics Journal 2019)  
[VERBALIZING REAL-TIME MATCH INSIGHTS AT THE US OPEN](#). [[US Open Article](#)] [[Times of India](#)] [[Digital Creed](#)]

### Conference

- **Jain, P.**, Lapata, M.,  
[Integrating Large Language Models with Graph-based Reasoning for Conversational Question Answering](#) (Under Review 2024)
- **Jain, P.**, Marzoca, A., Piccinno, F  
[Structsum Generation for Faster Text Comprehension](#) (ACL 2024)
- **Jain, P.**, Lapata, M.,  
[Conversational Semantic Parsing using Dynamic Context Graphs](#) (EMNLP 2023)
- Puduppully, R., **Jain, P.**, Chen, N., Steedman, M.,  
[Multi-Document Summarization with Centroid-Based Pretraining](#) (ACL 2023)
- Perez-Beltrachini, L., **Jain, P.**, Monti, E., Lapata, M.,  
[Semantic Parsing for Conversational Question Answering over Knowledge Graphs](#) (EACL 2023)
- Mittal, A., Saha, D, **Jain, P.**, Sen, J., Jammi., M.,  
[Bootstrapping Chatbot Interfaces to Databases Overnight](#) (CODS-COMAD 2021)

---

\* equal contribution

- **Jain, P.**, Mishra, A., Azad, A.P. and Sankaranarayanan, K.  
*Unsupervised Controllable Text Formalization* (AAAI 2019)
- Surya, S., Mishra, A., Laha, A., **Jain, P.**, Sankaranarayanan, K.  
*Unsupervised Neural Text Simplification* - (ACL 2019)
- Agrawal, P., **Jain, P.**, Dalmiya, A., Mittal, A.  
*Unified Semantic Parsing with Weak Supervision* (ACL 2019)
- **Jain, P.**, Laha, A., Sankaranarayanan, K., Nema, P., Khapra, M.M. and Shetty, S.,  
*A Mixed Hierarchical Attention based Encoder-Decoder Approach for Standard Table Summarization* (NAACL-HLT 2018, Short Paper)
- Nema, P.\*, **Jain, P.\***, Shetty, S., Laha, A., Sankaranarayanan, K. and Khapra, M.M.,  
*Generating Descriptions from Structured Data Using a Bifocal Attention Mechanism and Gated Orthogonalization* (NAACL-HLT 2018)

#### Workshop / Demo/ Preprint

- **Jain, P.**, Agrawal, P.  
Unified Semantic Parsing with Weak Supervision (EurNLP 2019)
- **Jain, P.**, Agrawal, P., Mishra, A., Sukhwani, M., Laha, A. and Sankaranarayanan, K.,  
*Story Generation from Sequence of Independent Short Descriptions* (SIGKDD Workshop ML4Creativity, 2017)
- Bharadwaj, S. *et al.* [and others, including **Jain, P.**]  
*Creation and Interaction with Large-scale Domain-Specific Knowledge Bases* (VLDB 2017, Demonstrations Track)
- **Jain, P.**, Balasubramanian, V.N.  
*Online Active Metric Learning for Clustering* (XRCI OPEN 2016)
- **Jain, P.**  
*Unsupervised Metric Learning Using Low Dimensional Embedding* (Preprint)
- **Jain, P.**  
*Metric Learning Tutorial* (Preprint)

## Selected Patents

- Mittal, A., Jain, P., Saha, D., Sen, J.  
*Ontology-Based Automatic Bootstrapping of State-Based Dialog Systems* - US Patent App. US20190228068A1
- Khare, S., Jain, P., Dechu, S., Mani, S., Tamilselvam, S.G  
*Enabling Tactile Senses in Digital Shopping* US Patent App. US20190163272A1
- Jain, P., Azad, A.P Mishra, A., Sankaranarayanan, K.  
*Unsupervised Tunable Stylized Text Transformation* - US Patent App. US20200034432A1
- Jain, P., Laha, A., Mishra, A., Sankaranarayanan, K.  
*Generation of Variable Natural Language Descriptions From Structured Data* US Patent App. US20200073944A1
- Sankaran, A., Laha, Lohia, P., Jain, P., Shrivastava, D., Agrawal, P.  
*Cognitive Assistant for Co-Generating Creative Content* - US Patent Application No. US20200134089A1
- Laha, A., Ekambaram, V., Jain, P., Shrivastava, D.  
*Displaying dynamic content on multiple devices* - US Patent Application No. US10664217B1
- Chemmengath, S., Jain, P., Laha, A., Krishnan, A.  
Script modification - US Patent Application No. US20200125638A1
- Jain, P., Laha, A., Mishra, A., Sankaranarayanan, K.  
*Structured Article Generation* - US Patent Application No. US11194816B2

## Talks

- Structsum Generation using Large Language Models - IIT Hyderabad, 2024
- Foster Talk at Jabalpur Engineering College, 2023
- Conversational Semantic Parsing - Online International. Conference. On. Advances in Physical, Mathematical and Computational Sciences, India - 2022
- Introduction to Machine Learning - Christ University Bangalore, 2018
- Metric Learning for Clustering in Streaming Large-Scale Data - IBM Bangalore, 2016
- Introduction to PageRank, IIT Hyderabad, 2015

## Awards

- Travel Award by Google to present our work at ACL 2024
- Travel Award by Facebook AI Research to present our work at EurNLP 2019
- IBM Second Invention Plateau Award - 2019
- IBM First Invention Plateau Award - 2019
- Equity Award for *outstanding achievement* in research and development. - 2018  
(An individual award selected globally every year by the office of CEO of IBM.)
- IBM Research First Patent Award - 2018
- Certificate of Appreciation from the Director of IBM Research India for my Contribution to Natural Language Generation Research in 2018
- Travel Grant by Xerox Research for attending XRCI Open, Bangalore - 2016
- Finalist Amazon Idea hackathon, Hyderabad 2016
- Travel Grant for attending the Fifth User Interaction Workshop at INCOIS, Hyderabad - 2015
- Ministry of Human Resource Development Scholarship by Government of India, 2013 - 2015
- Ministry of Human Resource Development Scholarship by Government of India, 2012

## Service

- Reviewer for COLM 2024
- I server regularly as a reviewer for ACL, EMNLP, NAACL. And have served in past for SIGKDD, Cods-Comad.
- Founding team member of TEDxIITHHyderabad, 2016
- Local coordinator for i.school workshop University of Tokyo, at IIT Hyderabad, 2015
- Student placement representative CSE Department IIT Hyderabad, 2015
- Core member of organizing team Aureoleâ€™10 Engineering Techfest, 2010

## Teaching Experience

### Tutor, University of Edinburgh

Jan 2021 - Apr 2021

Natural Language Understanding, Generation, and Machine Translation

### Teaching Assistant, IIT Hyderabad

July 2013 - June 2015

Numerical Linear Algebra,  
Database Management Systems,  
Advanced Compiler Design

## Mentorship

- Akash Kumar, Software Engineering intern at IBM (next PhD student at UCSD)  
Project: Tabular Data Summarization
- Aman Sinha, Software Engineering intern at IBM (next Research Assistant at IECL Laboratory, University of Lorraine)  
Project: Scheduled Multi-Task Learning for Semantic Parsing

## Relevant Coursework

- **PhD, Edinburgh** - Pragmatics, Controversies in the Data Society, Accelerated Natural Language Processing, Machine Learning and Pattern Recognition
- **M.Tech, IITH** - Hadoop, Advanced Algorithms and Data Structures, Mathematics for Computing, Systems Security

## Training and Certifications

- IBM Certified Academic Associate, -DB2 9 Database and Application Fundamentals (ID: PR1587378)
- IBM Certified Deployment professional, -Tivoli Directory Server V6.3 (ID: PR1587378)
- Training on Computer Networking, -CMC Academy, Jabalpur (45 days)